# Regional Disparity and Dynamic Development of China: a Multidimensional Index

Peng Bin

School of International Studies - University of Trento

February 2015

# Regional Disparity and Dynamic Development of China: a Multidimensional Index

*SIS Working Paper N^o^* 2015–1

February 2015

Peng BIN
*University of Trento*

School of International Studies
via Tommaso Gar, 14
38122 Trento – ITALY
http://www.unitn.it/ssi

## Regional Disparity and Dynamic Development of China: a Multidimensional Index

Peng Bin[1]

**Abstract:** This paper investigates the evolution of regional socio-economic disparity in China during the period 1998-2010. A new composite index of development (CIRD) is developed to capture the five main dimensions of regional development: macroeconomic conditions, science and innovation performance, environmental sustainability, human capital accumulation, and public facilities provision. The investigation benchmarks 30 (out of 31) Chinese regions according to such multidimensional index of development and thus improves the understanding of the evolution of regional disparity in China in terms of the various dimensions of socio-economic development. Finally, on the basis of stochastic kernel density estimation, the paper reveals the existence of a triple-clubs pattern of convergence in the period under scrutiny, thereby informing both the literature on regional convergence and the current strategy of balancing the uneven process of growth in China.

**Key words:** regional disparity, multidimensional index of development, stochastic kernel, distribution dynamics, Chinese regions

**JEL codes:** O18, P48, R58

[1] Post-doc fellow, School of International Studies, University of Trento, Via Tommaso Gar, 14 - 38122 Trento, Italy.
Telephone: +390461283132, Email: peng.bin@unitn.it

## 1. Introduction

Regional disparity has always been a hot issue in China as the country has been struggling to follow a balanced growth path and is still confronting with unprecedented pressures and challenges. An interesting representation of the regional disparity in the country has been offered in "One China, Four worlds" (Hu, 2003)[2], according to which the Chinese regions can be classified into four groups according to their income level: the first class is composed by the most developed regions (such as Shanghai and Beijing); the second class consists in the majority of East coastal provinces and municipalities (such as Tianjin, Guangdong, Zhejiang, Jiangsu, Fujian); the third class is composed by the North-East and some Central regions (such as Hebei and Shanxi); and the fourth class includes the peripheral regions (such as Guizhou and Yunnan). The income gap between the first and the fourth classes was and still is enormous, as the regions in the latter group have lagged behind for half a century and still are far from the economic development level of the group ahead.

The complexity of Chinese regional disparity is hard to depict without paving the policy background. The East coast has certainly benefited a head-start from the decentralization of Chinese regional policies in the early stage of its marketization reform from the 1978 onwards. In the framework of Deng Xiaoping's guideline for prioritizing the development of certain regions[3], the Special Economic Zones (SEZ) were designed and 14 coastal cities were open to foreign investors in 1984. This approach had a twofold result. On the one hand, it facilitated market liberalization and boosted its exporting sectors by attracting Foreign Direct Investments (FDI) (Bonatti and Fracasso, 2010). On the other hand, it fostered uneven development and regional inequality as a side effect. The Chinese central government realized the unintended consequences of its geographically-biased policies and commenced a series of preferential policies to support rural and inland regions since the late 1990s, aiming to balance the previous disproportional benefits reserved to the coastal regions. In 1996, the "Ninth Five-year Plan" was launched, aiming to gradually narrow the existing large regional gaps. A number of supporting policies and programs were developed. For instance, the 15th Party Congress in 1997 emphasized the need to improve the structure of regional development and to coordinate growth policies across regions. Accordingly, in the "Tenth Five-year Plan" started in 2001, a new regional program of "Western Development (Go-West)" was initiated. This was followed by several other regional programs implemented in 2003 and 2004, respectively targeting at the revitalization of the traditional Northeast industry and at the promotion of central China ("Go-Central"). The "Eleventh Five-year Plan" followed the guideline of a coordinated and balanced growth. It aims to improve the mechanism of regional coordination, thus to promote the common prosperity of the Eastern, Central and

[2] This representation was based on the World Bank's criterion of income group, with the application of GDP per capita ($ PPP).

[3] Deng Xiaoping delivered the speech that it should be allowable for some people and some regions to get rich before others for the very first time in 1985, and he recurred to this subject twice in 1986.

Western areas. In general, the development strategy followed by the central government exhibited a gradual shift of the attention from the Eastern regions to the Central and Western areas of the country. This implies a pattern of growth characterized by the emergence of the growth clubs, where the development of the coastal regions was the first priority, and inland regions had to catch up later.

Notwithstanding these governmental measures to balance the uneven growth process, regional disparity still remains high: According to the study of Fan, et al. (2011), the rural-urban income disparity has accounted for a large proportion of the total income inequality increase in the country during the period 1952-2008, and this rural-urban gap of income per capita is even wider in the inland regions[4]. But the income gap between the coastal and the inland has also widened up starting with the late 1970s. Such heterogeneous growth process is reflected at the micro level across households and individuals: the Gini coefficient of income inequality in China has remained in the range of 0.473 and 0.491 during the last decade (2003-2013), which are very high values if compared with the international alert line[5].

Regional disparity can be observed in other aspects besides the income gap. Not by chance, many strategies and investment plans have been developed by the authorities to improve the general living conditions and to protect the natural environment in the underdeveloped Western regions and rural areas, given that the Eastern region still exhibits a relatively more advanced level in education, public health, transportation, employment, and the like[6]. As regional development regards the economic, social and environmental realms, both scholars and policy-makers do not typically focus exclusively on income disparity and extend their interest on the multi-dimensional aspects of regional inequality in China. Given the political importance of a coordinated development process and considering the actual high levels of regional disparity, this topic is particularly important for the Chinese authorities. In the past, the GDP has been commonly used as a synthetic indicator of the level of well-being in China and, accordingly, the distribution of GDP per capita was used to assess the regional disparities in the level of development. The GDP per capita, however, is not a variable capable of capturing the entire socio-economic picture. The very Chinese authorities have recently acknowledged the limitations of focusing on GDP per capita as a single measure to evaluate the performance of local

---

[4] According to their calculations, urban income per capita was almost threefold of rural in the coastal regions, and 3.2 times in the inland regions in 2007.

[5] The international alert line (0.4) is set as a threshold beyond which inequality might become a concern for social stability by UN-Habitat (2008). The mainland authority of China, however, adjusts the “alert line for wealth gap” at 0.45. In January 2013, China has published its Gini coefficients of the past decade (2003-2012) for the very first time in 12 years. Its Gini coefficient has reached the lowest (0.473) in 2004 and peaked at the highest (0.491) in 2008; it has been decreasing since 2008, and reached the lowest level (0.474) of the past seven years in 2012.

[6] The Eastern region occupies only 9.5% of the national land, whereas it accounts for 38.2% of total population and 51.3% of national GDP. According to the latest Chinese statistical annual report published by National Bureau of Statistics in 2013, 39.1% of total number of regular institutions of higher education locate in the Eastern region, and its distribution density of regular institutions of higher education is 10.4 unit per 10,000 sq.km, whereas the Western region is only 0.9 unit per 10,000 sq.km; 35% of hospitals locate in the East, with relatively large shares of licensed doctors (41%) and hospital beds (38.1%) in national total, and its distribution density of hospitals is 88.44 unit per 10,000 sq.km, whereas the West is only 10.5 unit per 10,000 sq.km; the East also takes large shares in the total volume of passenger and freight in transportation (38.9% and 52.3% respectively); and it has the lowest registered unemployment rate in urban areas (3%).

leaders, and they started attaching greater attention to the reduction of inequality, as well as to the promotion of a widespread social development and the protection of the environment[7]. Therefore, developing and investigating new synthetic indexes and measures that capture the multifaceted dimensions of socio-economic development is of utmost importance for both the researchers and the Chinese policymakers. Such an approach allows to reach a more comprehensive understanding of the diversified regional development levels and of the evolution of regional disparities in China over time. Notwithstanding its importance, very little research and empirical evidence has been produced on this issue and this work aims at filling this gap in the literature.

The creation of synthetic indexes is a complicated undertaking due to scarce availability of data and various technical aspects. This work tackles such problems and develops a new *Composite Index of Regional Development* (CIRD), involving five different sub-dimensions of the concept of development, including *macroeconomic conditions*, *science and innovation*, *human capital*, *environmental sustainability* and *public facility*. The development of the CIRD at a relatively highly disaggregated level and its calculation over a considerable time span (i.e., 1998-2010) are among the major contributions of this study to the literature on regional inequality in China. The CIRD index, and some sub-pillar indexes, allows observing the dynamics of the multidimensional process of development across the Chinese regions over the interval 1998-2010, which embraces three "Five-year Plans"[8]. It is the very first time adopts a composite index to evaluate dynamics of distribution across 30 (of 31) regions, in the attempt to overcome the limitations of using a solitary measure of GDP or income per capita as done in previous studies.

The remaining of the paper continues as follows: *Section 2* sketches the methodologies and the data in this study; *Section 3* is dedicated to estimate the regional disparity and dynamics of distribution according to the CIRD in China. *Section 4* draws some closing remarks.

## 2. Research design: methodology and data

This paper is based on, and adds to, two blocks of the methodologies: the first one regards composite indexes of development, and the second block refers to the studies of dynamic distribution based on stochastic kernel density estimation and regional disparity in China. The data come from Chinese statistical yearbooks published by the National Bureau of Statistics of China to guarantee reliability and consistency. This study builds up a large panel dataset involving 31 Chinese provinces and municipalities from 1998 to 2010. Tibet is removed in two sub-pillars (SII and ESI), and thus also in the CIRD, because of missing data. It is worth noticing that, given the existing problems in

[7] In the national conference on organizational work of 2013 in China, President Xi Jinping underlined that "we should no longer evaluate the performance of leaders simply by GDP growth. Instead, we should look at welfare improvement, social development and environmental indicators to evaluate leaders". Until now, more than 70 Chinese prefectures have abandoned the GDP as the sole indicator to assess the leaders' performances and local government records.

[8] Given the missing data for various aspects of interest in the Chinese yearbooks for 1996 and 1997, and considering that the Chinese yearbook changed structure and calendar in 1998, this study choses to focus on the time interval 1998-2010.

matching Chinese datasets, the construction of such large a dataset at the provincial level represents in itself a contribution to the literature on the Chinese regional development.

*The 1st block: the construction of CIRD based on PCA*

All the numerous studies that examine different types of indicators to construct indexes of socio-economic development are based on the tenet that a limited group of indicators cannot adequately capture the various aspects of development. A vast literature has attempted to measure regional development conditions through composite indexes. One of the most popular indexes is the *Human Development Index* (HDI)[9], produced and reported by the United Nations Development Programme (UNDP). Less known but equally interesting indexes are those developed by either individual researchers or institutions to map the various dimensions of regional development in, such as the *Basic Capabilities Index* (BCI, previously *Quality of Life Index*)[10], the *Capital Access Index* (CAI)[11], the *Commitment to Development Index* (CDI)[12], the *Country Performance Assessment* (CPA)[13], the *Economic Freedom of the Word Index* (EFW)[14], the *Education for all Development Index* (EDI)[15], *the Global Competitiveness Index* (GCI)[16],. Environmental sustainability has increasingly become a significant subject of research and composite indexes have been developed too: the OECD environmental performance review, the *Environmental Sustainability Index* [17] (ESI), and the *Environmental Vulnerability Index* (EVI)[18]. In the case of China, environmental sustainability has gained importance and it has been integrated into the *China Urban Sustainability Index* (UCI, 2010)[19].

[9] HDI is based on three main components: life expectancy, adult literacy rate and GDP per capita.

[10] The BCI depicts basic well-being via an efficient rating for the basic levels of people's well-being according to their health state and performance in primary education. It was created by Social Watch, initially named as Quality of Life Index in 2004 and renamed in 2005. It is reported annually and with coverage of 94 countries.

[11] The CAI aims to evaluate the ability of entrepreneurs to gain access to financial capital in 85 countries around the world, including five dimensions of capital access: macroeconomic and institutional environment, banking system, financial market development, international access to funds and sovereign ratings, descried by 54 variables from 15 different data sources. It was created by Milken Institute in 1985 and reported annually.

[12] The CDI reports the dedications of 21 OECD countries to policies that benefit the 5 billion people living in poorer nations, through seven categories: quality of foreign aid, openness to developing country exports, policies of investment, migration policies, support for creation of new technologies, security policies, and environmental policies. It was created by Center for Global Development and Foreign Policy in 2003 and reported annually.

[13] The CPA focuses on policies and institutional arrangements of 26 Asia Pacific countries, through seventeen indicators to describe coherence of macroeconomic and structural policies, policy promotion of equity and inclusion, governance and management quality and portfolio quality. It was created by Asian Development Bank in 2006 and reported annually.

[14] The EFW estimates the degree of economic freedom of 123 countries, involving five aspects: size of government, legal structure and protection of property rights, access to sound money, international exchange and regulation. It was created by J. Gwartney and R. Lawson from Fraser Institute in 1996 and reported annually.

[15] The EDI measures the progress towards the goals of Education for All (EFA). It describes four goals: universal primary education, adult literacy, quality of education and gender. It is created by UNESCO in 2002 and reported annually, covering 127 countries.

[16] The GCI was launched in the 2004/2005 Global Competitiveness Report. It includes three sub-pillars as basic requirements, efficiency enhancers and innovation factors. It was created by World Economic Forum (WEF) and reported annually.

[17] The ESI was created by Center for International Earth Science Information Network (Columbia University) and Yale Center for Environmental Law and Policy (Yale University) in 2000, covering 146 countries. It reflects the progresses of environment sustainability in five dimensions: environmental systems, reducing stresses, reducing human vulnerability, social and institutional capacity and global organization.

[18] The EVI was created by South Pacific Applied Geoscience Commission (SOPAC) in 2004, covering 23 countries.

[19] The *China Urban Sustainability Index* was built up in 2010 by the *Urban China Initiative* and reported annually. It has five categories: basic needs, resource efficiency, environmental health, built environment, and commitment to sustainability.

The aggregation of various indicators covering diverse socio-economic realms into a single index has been identified as a very promising solution. Parametric approaches to aggregate indicator have been widely accepted as the most suitable ways to define the weights for the indicators entering a composite index. One of the most common methodologies is the *Principal Component Analysis* (PCA). It is a data reduction method introduced by Hotelling (1933) that allows identifying few unobservable factors that account for most of the variability in an underlying large set of indicators. This method is frequently applied to integrate different indicators or sub-indexes, since it permits a relatively easier selection of the sub-set of components of interest (Mazzocchi, 2008) and serves a simpler application without assumptions attached to the original data (Chen and Woo, 2010). There will be a long list if we numerate the literature using PCA to aggregate a composite index or a multidimensional measurement. Dreher (2006) and Heshmati (2006), for example, applied PCA to estimate the weights of sub-indexes of a composite globalization index. Chen and Woo (2010) used PCA to obtain a composite index of economic integration in the Asia-Pacific region. The premise to run a PCA is to use the original variables that are highly correlated, and to identify a reduced number of uncorrelated linear combinations from these initial correlated variables. It also requires to subtract the mean from each of the data dimensions to run the PCA properly[20] (Smith, 2002). The key of PCA is to maximize the variability explained by the components: the *principal components* are indeed those describing most of the variability.

This study builds up a multidimensional dataset $\mathcal{X}_{p} = (x_1, x_2, \dots, x_p)$ for the 30 observations (regions) in each year, where $p$ represents the multiple dimensions (indicators) to constitute the composite index (CIRD). Then the PC can be expressed as:

$$\mathcal{C} = \mathcal{X}\mathcal{A}' = \begin{cases} c_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p \\ c_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p \\ \qquad \cdots \\ c_p = a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pp}x_p \end{cases} \tag{3.11}$$

here $\mathcal{C}$ represents the *principal components* as a series of linear combinations of the indicators $\mathcal{X}_{p} = (x_1, x_2, \dots, x_p)$; $\mathcal{A}'$ is the eigenvector matrix where $a_{ip}$ $(i = 1, 2, \dots, p)$ represents the $i^{th}$ eigenvector of the indicator correlation matrix.

To re-aggregate the *principal components* is the subsequent step of the analysis and it requires great attention. This study determines the contribution to the CIRD by using a parametric method with assigned weights. These latter are calculated by the following equation:

$$\mathcal{W}_j = \frac{\sum_{i=1}^{i=p} \lambda_i a_i}{\sum_{i=1}^{i=p} \lambda_i} \tag{3.12}$$

here $\lambda_i$ is the $i^{th}$ eigenvalue of the indicator correlation matrix, which denotes the variance of the $i^{th}$ *principal*

[20] The formula here is $\bar{x}_i = \frac{x_i - x_{mean}}{st.deviation(x_i)}$, here $x_i$ is the values of $i$ observations (regions) for each indicator.

*component*[21]. In order to overcome the bias weights that might be caused by the highly-correlated indicators (Mishra, 2007), this study uses a two-stage PCA recommended by Chen and Woo (2010)[22]: firstly the PCA is run for the five sub-dimensional indexes in every year of the sample, and obtain plural *principal component(s)*[23]; then the PCA is proceeded again with these plural *principal component(s)* in every year, and construct the final composite index (CIRD) by the assigned weights by the formula *(3.2)*. According to the Kaiser rule, this study always selects the *principal component(s)* with eigenvalues higher than 1.

This works embraces the PCA approach to develop an innovative composite index – the CIRD – able to capture a comprehensive measure of regional development in China. Though not new in terms of methodology, the CIRD is new as it qualifies, to the best of our knowledge, as the first composite index able to capture five main dimensions of the regional development process in China.

*The 2nd block: dynamic distribution and regional disparity in China*

The stochastic kernel estimation to address regional dynamics was first developed by Quah (1995, 1996, and 1997). Quah established a new framework based on the *intra-distribution dynamics*, which allowed him to observe a profound empirical regularity of "*emerging twin peaks*" in the cross-country income distributions over time. This non-parametric method is based on Markov transition probability matrix to estimate the transition process from one state to another.

This method has been applied in a relatively few works to describe the transitional dynamics of convergence in China. These studies focus on different administrative unit levels and time intervals. Most studies hold a commonly-accepted view that a bi-modal or twin-peak pattern has emerged in the distribution dynamics of regional convergence in China, yet they differ along various dimensions. Some works focus on the provincial level with long-term coverage (Li, 2003; Zou and Zhou, 2007; Sakamoto and Islam, 2008; Villaverde, et al., 2010; Barone, et al., 2013). Li (2003) found little possibility for the Chinese provinces to converge in real GDP per capita during 1978-1998, a result supported also by the investigation of Sakamoto and Islam (2008) who, looking at the period of 1952-2003, concluded that long-run convergence remained as an open issue. Zou and Zhou (2007), Villaverde, et al. (2010) and Barone, et al. (2013), instead, provided evidence of a pattern of growth club. Other works focused on the lower administrative unit level in special, restricted zones (Zhou and Zou, 2010; Wei and Ye, 2009; Sakamoto and Fan, 2010; Liu and Zou, 2011; Cheong

[21] The eigenvalues corresponding to each component represents the amount of variance they explain; the sum of eigenvalues equals the original number of variables, therefore a lower level of eigenvalues explains less than the standardized variance which should be equal to 1.

[22] Considering the properties of our data and index, this study makes some adjustments in running the two-stage PCA. Instead of "grouping the highly inter-correlated indicators together and construct a composite sub-index first" (Chen and Woo, 2008, 2010), this study runs the first-stage PCA in the five sub-indexes, since the indicators among different sub-indexes are not very highly-correlated.

[23] There are 6 PCs in the years of 1998, 2006, 2007, 2009 and 2010 respectively and 5 PCs in the years of 1999-2005 and 2008 respectively.

and Wu, 2013). Wei and Ye (2009), for instance, used prefectural within the Zhejiang province and found a polarization between the coastal and interior Zhejiang. Sakamoto and Fan (2010) covered 75 cities and counties in Yangtze River Delta and revealed a bi-modal distribution dynamics of the per capita income during 1990 to 2005. Zhou and Zou (2010) concluded about the existence of twin peaks in the dynamic distribution of urban income by focusing on prefectural cities during the period 1995-2004. Liu and Zou (2011) investigated an unimodal pattern and a twin-peak pattern of the income dynamics in rural and urban China respectively. Cheong and Wu (2013) applied county-level data from 1997 to 2007 and claimed a persistence pattern in the Chinese spatial groups.

As anticipated, the first common limitation of these studies is that they mainly rely on GDP or income per capita as their sole indicator to assess convergence. In line with what written in the introduction, however, this is a relatively narrow approach that prevents from capturing many other socio-economic dimensions of development, all important from a theoretical and political viewpoint. This work aims at making progress in this direction and this is why a new composite index will be used to map the evolution of regional disparity in China.

This study, as anticipated, does not intend merely to illustrate the persistent differences across the regions. Rather, It develops a stochastic kernel density estimation to depict the dynamic trajectory of regional disparity in China, by focusing on a comprehensive measure of development. Following Quah (1993, 1995, 1996, 1997), this study adopts the stochastic kernel density to tackle the dynamic distribution of regional development in China. The method takes the Markov transition probability matrix from one state to another as a version of a function of probability density defined on current and future levels. It allows identifying the formation of convergence clubs, polarisation or persistence inequality, as well as other patterns. Let assume that the distribution of the CIRD scores at time *t* is $d_t$, then the Markov chain can be used to describe the dynamics of this distribution at time *t*:

$$d_t = T(d_{t-1}, u_t) = T_{u_t}(d_{t-1}), \;\; t \geq 1 \qquad (3.21)$$

where $u_t$ is the disturbance term, $T$ is an operator that describes the distribution transition, $T_{u_t}$ is the operator absorbing the disturbance term. By iterating this process using a Markov chain assumption, one can obtain the long-run distribution:

$$\mathcal{D}_{t+s} = (M^s)' \, \mathcal{D}_t\,, \;\; s \geq 1 \; and \; s \rightarrow \propto \qquad (3.22)$$

where $\mathcal{D}_{t+s}$ is the long-run distribution of the CIRD scores across regions after $s$ period, and $M$ is a complete description of distribution transitions from the state of the CIRD at time $t$ to the state at time $t + s$.

It is useful to illustrate in a nutshell how to interpret these values. When the distribution $\mathcal{D}_{t+s}$ tends to move towards one point mass, this can be considered as a sign of convergence. On the contrary, polarisation or stratification are associated with $\mathcal{D}_{t+s}$ that tend to be bimodally or multimodality distributed. According to Quah (1997), the stochastic kernel can be defined as follows:

$$f(x_{t+s}) = \int \mathcal{M}\,(x_{t+s}|x_t)f(x_t)dx \tag{3.23}$$

where $x_t$ and $x_{t+s}$ are the initial level of the CIRD and the CIRD level at time $t+s$, respectively; with the assumption that $\{(x_t^1, x_{t+s}^1), (x_t^2, x_{t+s}^2), \ldots, (x_t^{30}, x_{t+s}^{30})\}$ denotes the set of the CIRD scores across 30 Chinese regions, $f(x_t)$ and $f(x_{t+s})$ are the density functions denoting the CIRD distribution cross regions at initial time $t$ and time $t+s$, respectively; $\mathcal{M}(x_{t+s}|x_t)$ represents the stochastic kernel mapping the conditional distribution after time $s$.

This approach identifies the distribution dynamics by applying the main concepts of stochastic kernel. To operationalize these concepts, the kernel density estimation is a key step. The estimator $\tilde{f}(x)$ calculating the percentage of regions close to the point $x$ can be defined as (Zambom and Dias, 2013):

$$\tilde{f}(x) = \frac{1}{nh}\sum_{i=1}^{n}\mathcal{K}\left(\frac{x-x_i}{h}\right) \tag{3.24}$$

where $\hbar$ is the bandwidth to measure the smoothness of the distribution curve at a given moment in time; $x_i$ is one random point to gauge the density distribution by its distance to $x$; $\mathcal{K}(\cdot)$ is the kernel function with symmetry as $\int \mathcal{K}(x)dx = 1$. The kernel density estimation is significantly determined by the appropriate selections of the kernel function and the bandwidth. There are several kernel functions: Biweight, Cosine, Epanechnikov (alternative: Epan2), Gaussian, Parzen, Rectangular and Triangular. The Epanechnikov function is the default function with high efficiency in minimizing the mean integrated squared error. Considering this study only has 30 observations, this study uses the alternative Epan2 function, which can be expressed as follows (Hardle, 1990; Cameron and Trivedi, 2005):

$$\mathcal{K}(u) = \begin{cases} \frac{3}{4}\left(1-\frac{1}{5}u^2\right) & if\ |u| < 1 \\ 0 & otherwise \end{cases} \tag{3.25}$$

The bandwidth $\hbar$, as another key factor to drive the number of values that are included in the density estimation at every single point, can be determined as:

$$\hbar = 0.9 * n^{(-\frac{1}{5})} * min\left(\sqrt{variance(x)}, \frac{interquartile\ range\ (x)}{1.349}\right) \tag{3.26}$$

where $x$ is the CIRD values to estimate the kernel and $n$ is the number of regions which is 30 in this study.

## 3. Regional disparity and dynamic transitions based on the multidimensional index

In this section, this study develops a stochastic kernel density estimation to depict the dynamic trajectory of regional disparity in China. As mentioned, this study is interested in a more comprehensive measure of development and, moreover, it aims to map the trajectory of regional disparity in China over the 13 years.

### *3.1 The Composite Index of Regional Development (CIRD)*

In this work, in particular, it is important to include variables capturing the various aspects of socio-economic development. It goes without saying that the selection of appropriate indicators that underpin any aggregate index requires a comprehensive understanding of the issues at stake. The indicator selection follows a number of criteria that an index should conform with. The first criterion regards the ability of the index to take a full-size snapshot of the socio-economic development of the regions; it must cover various aspects of development. The second criterion is that the index must be capable to track down the evolving regional disparity (in China), hence it must be based on the individual indicators that are consistent and comparable both horizontally (among regions) and vertically (among years). A number of other conditions must be equally satisfied: the indicator selection must be underpinned by a solid theoretical framework; the indicators must be available for data reduction and transparent to be applied by other studies, and etc. (Saisana and Tarantola, 2002; OECD, 2008). Endowed with a  good composite index , one can answer such questions as what regions are leading and lagging behind in terms of socio-economic development, what indicators contribute more to determine the level of local development, and the like. This makes the index a powerful tool both for comparative purposes and for policy goals.

This study selects a large set of available indicators at the Chinese provincial level over the time period 1998-2010, in the attempt to cover multiple domains of the regional socio-economic development in China (such as production, investment, trade, employment, human resources, infrastructures, environment protection, science and technology ). These indicators, which are described in Table 1, are grouped into five sub-indexes which represent what can be identified as the main pillars regional development in China: *a Macroeconomic Index* (MEI)*, a Science and Innovation Index* (SII)*, an Environmental Sustainability Index* (ESI)*, a Human Capital Index* (HCI) *and a Public Facility Index* (PFI) [24].

[24] It is worth noticing that to identify the 25 indicators constituting the CIRD, a larger dataset of potential indicators has been built and analysed. The 25 indicators (5 in each sub pillar) were selected as following: first, the non-significant indicators are removed wisely by their poor correlations with the GDP growth rate; then the PCA analysis was applied within each pillar. The choice of the five baseline indicators for each pillar has been made by observing what variables accounted for most of part the variability in the principal components identified for each dimension.

***Table 1.*** *The indicators of five-dimensional CIRD*

| | |
|---|---|
| **Macroeconomic Index (MEI)** | |
| GRP per capita | GRP per capita (deflated) (10,000 Yuan) |
| FDI | proportion of foreign direct investment in GRP (%) |
| trade balance | trade ratio of export-import (ratio) |
| consumption | final consumption rate (%) |
| compensation | employ compensation proportion over GRP (%) |
| **Science and Innovation Index (SII)** | |
| government expenditure SII | proportion of government expenditure in science and innovation over total government expenditure (%) |
| labor productivity in high S&T firms | average labor productivity of high science and technology enterprises (deflated) (10,000 Yuan per worker) |
| income per capita in high S&T firms | average income per capita in high science and technology enterprises (deflated) (10,000 Yuan per worker) |
| R&D expenditure level | R&D expenditure proportion over GRP (%) |
| trade level in S&T market | trading value in science and technology market over GRP (%) |
| **Environmental Sustainability Index (ESI)** | |
| industry waste water cleaned | proportion of discharge of industrial waste water meeting standards (%) |
| utilization industry waste | utilized ratio of industrial waste (%) |
| output from waste | output value of products made from waste gas, water and solid wastes per worker (deflated) (10,000 Yuan) |
| living waste | volume of living waste discharged per capita (ton) |
| industry anti-pollution | investment of industrial enterprises on anti-pollution projects per worker (deflated) (10,000 Yuan) |
| **Human Capital Index (HCI)** | |
| educated proportion | proportion of educated population (%) |
| employ 3rd sector | employ ratio of 3rd sector over 1st sector (%) |
| working age distribution | working age population distribution (%) |
| life expectancy | life expectancy (estimated by the population census)[25] |
| urban proportion | urban population proportion (%) |
| **Public Facility Index (PFI)** | |
| government expenditure in public service | governmental expenditure proportion in public services over total government expenditure (%) |
| public vehicles | public transportation vehicles per 10,000 persons (unit) |
| paved road | area of paved road per 10,000 persons (sq.m.) |
| education | public schools per 10,000 persons (unit) |
| medical health | public medical services per 10,000 persons (unit) |

*Source: author's own collection and elaboration from various Chinese statistical yearbooks*

The 25 indicators in these five groups enter the CIRD according to weights found via the methodology presented in *Section 2*. Table 2 reports the rankings of these 30 Chinese regions by the average scores and the average changing levels of the CIRD during the 13 years of the sample[26]. It emerges a tendency of stratification since there are clear gaps among the average CIRD scores. In contrast, the ranking grades in the right hand, which denote the average changes during the 13 years calculated by the average CIRD scores of every region, present a smooth decline. The bulk of the fast-growing regions (the top one-third in the right-hand ranking, except Beijing, Zhejiang and Shanghai) are graded

[25] The life expectancy is calculated by the national population census reports 1990, 2000 and 2010.

[26] Min-Max is used here to make the ranking reports reader-friendly.

The Min-Max formula is expressed as: $z = \frac{(x_i - x_{min})}{(x_{max} - x_{min})} * 100$

below average in the left hand. More than half of the regions present negative growth of the index[27]; however, all the variations, no matter positive or negative, are within a narrow range (Appendix 3 shows the original data). Given that this work uses the mean-subtracted data, the moderate variations actually indicate a tendency of persistence across Chinese regions in their socio-economic development process.

***Table 2. The Rankings of Chinese regions in terms of the Min-Max scores of CIRD (1998-2010)***

| Regions | Average CIRD | Regions | Average changing levels |
|---|---|---|---|
| Beijing | 100 | Beijing | 100 |
| Shanghai | 87 | Zhejiang | 99 |
| Jiangsu | 62 | Ningxia | 97 |
| Shandong | 58 | Jiangxi | 92 |
| Zhejiang | 57 | Shaanxi | 87 |
| Tianjin | 57 | Inner Mongolia | 82 |
| Guangdong | 54 | Shanghai | 80 |
| Fujian | 38 | Shanxi | 80 |
| Liaoning | 38 | Guizhou | 80 |
| Hubei | 31 | Anhui | 79 |
| Hebei | 25 | Chongqing | 75 |
| Henan | 24 | Hubei | 69 |
| Heilongjiang | 23 | Yunnan | 67 |
| Hunan | 22 | Henan | (63) |
| Hainan | 22 | Guangxi | (61) |
| Chongqing | 21 | Hebei | (59) |
| Jilin | 21 | Guangdong | (58) |
| Shanxi | 21 | Gansu | (54) |
| Anhui | 20 | Shandong | (47) |
| Sichuan | 19 | Qinghai | (47) |
| Shaanxi | 19 | Hunan | (45) |
| Yunnan | 14 | Xinjiang | (44) |
| Jiangxi | 14 | Tianjin | (39) |
| Guangxi | 13 | Fujian | (36) |
| Inner Mongolia | 10 | Sichuan | (29) |
| Xinjiang | 8 | Jiangsu | (25) |
| Gansu | 8 | Jilin | (19) |
| Ningxia | 6 | Hainan | (17) |
| Guizhou | 3 | Liaoning | (16) |
| Qinghai | 0 | Heilongjiang | (0) |

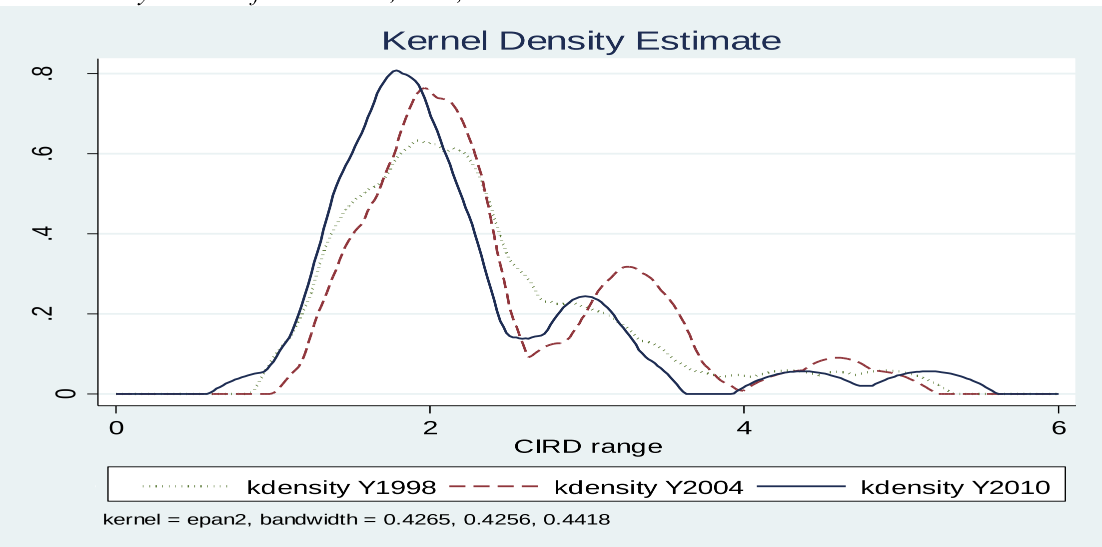
### 3.2 *Distribution dynamics of regional socio-economic development: a stochastic kernel estimation*

Based on the concepts discussed of stochastic kernel discussed in *Section 2*, this study estimates a persistent stratification on socio-economic level across 30 Chinese regions during 1998 to 2010. Figure 1 reports three kernel density curves of the CIRD estimated for the years 1998, 2004 and 2010. One can observe that the marginal distribution remains a pattern of multimodality in the three years, and this multimodality of the distribution becomes more clear over time. At the beginning (1998), multimodality is not patent. The kernel density curve shows a main mode indicating

[27] The regions with negative growth are presented with the scores in parentheses. Appendix 1 shows the negative growing regions in the map (the regional names are presented in Appendix 2).

that around 60% of Chinese regions fall between 1.75 and 2.25 times of the average CIRD, and a long right tail indicating that the rest appears to distribute more uniformly in the levels higher than 2.25 times of average CIRD. In the middle year (2004), the kernel curve shows a clear pattern of triple modality. There is a small but clear increasing tendency in the main mode of the distribution compared to the 1998's main mode. The previous uniform distribution in the higher levels breaks into two modes: one gathers around 3.25 times the average and one gathers around 4.5 times the average. In the last year (2010), the multimodality pattern remains and moves slightly leftward, as a result of a small mobility to the lower level groups. These three kernel curves suggest that a stratification remains across Chinese regions in their distribution based on socio-economic development, and that the Chinese regions tend to converge (eventually) towards several different values of the CIRD, in line with a sort of club-convergence.

***Figure 1.*** *Kernel Density Curves of CIRD 1998, 2004, 2010*




*Source: the data are collected from the Chinese Statistical Yearbook and calculated by the authors;*
*Note: the x-axis is benchmarked with the national average level of CIRD; the y-axis is the distribution density.*

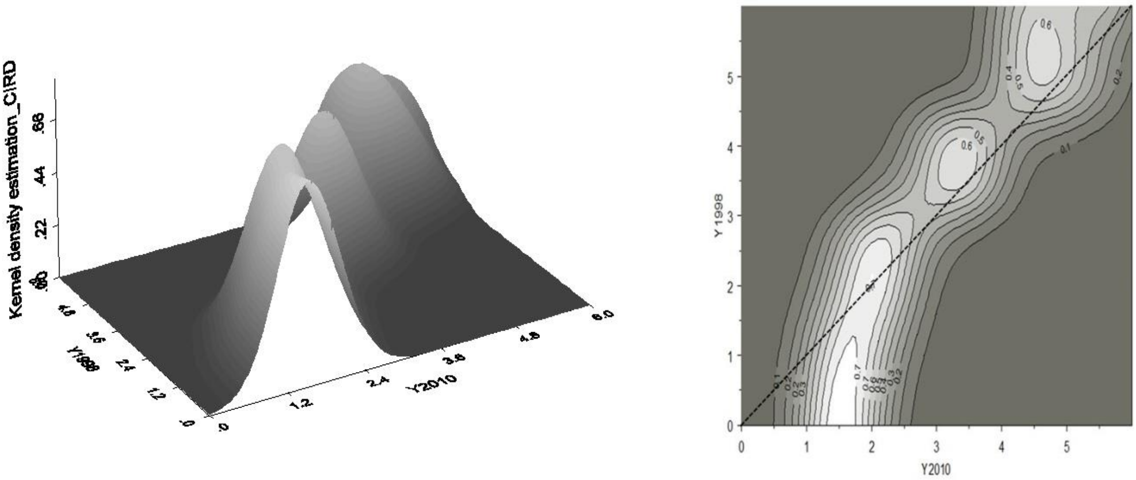
In order to investigate the dynamic distribution of socio-economic development across Chinese regions, this work then uses a three-dimensional plot and a contour plot of the stochastic kernel for the 30 Chinese regions (Figure 2). Both reveal a pattern of triple-peak distribution in China and a persistence of the dynamic distribution of the socio-economic development levels among the Chinese regions. In the 3D plot, the x-axis represents the relative distribution of the CIRD for the last year 2010, while the y-axis represents the same for the first year 1998; the z-axis represents the stochastic kernel[28]. The heights of the peaks indicate the activeness of the transitions during the 13 years: the higher the peaks, the more dynamic the transitions of the distributions in the corresponding parts. The contour plot is the corresponding projection of the 3D kernel surface in the two dimensions, where each line connects all the points on the stochastic kernel with the same height.

One can clearly observe there are three groups gathering around 1.5, 3.5 and 5 values on the x-axis. A strong

[28] It means how the conditional density with which a part of distribution in 1998 ends up as another part of distribution corresponding to 2010.

persistence of these values is also revealed by the fact that all the three modes (or peaks) lie on the diagonal. This implies that there is a high probability that a region stays in the same or similar position of the distribution of regional socio-economic development during the 13-year period. To be precise, however, there is also a tendency of counter-clockwise movement with respect to the diagonal, especially if one inspects the first group. Such pattern indicates that some members in this group developed fast and drove the whole group towards the middle one. Hence, the stochastic kernel plot provide evidence of a stratification and of convergence club formation in the socio-economic development of China during 1998-2010, stemming from the formation of a middle level group and the fast growth of some members in the low level group.

***Figure 2.*** *Three-dimensional Kernel Surface and Contour Plot of CIRD 1998-2010*



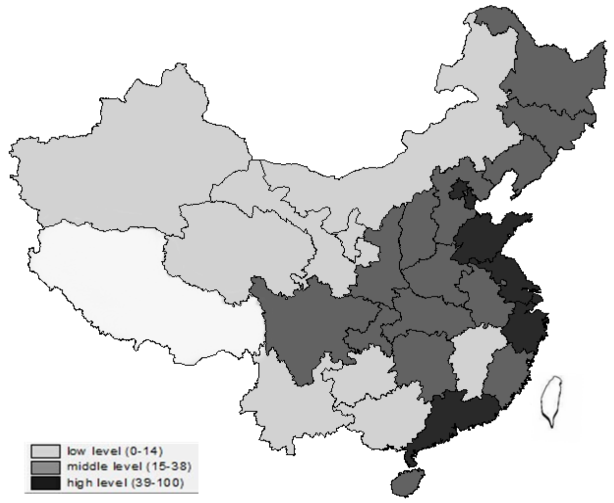
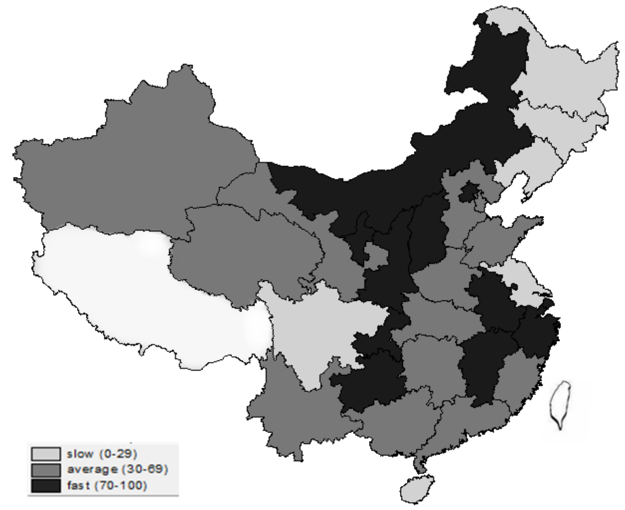

*Source: this figure is got from the dataset collected and calculated from the China Statistical Yearbooks 1999-2011 by the author, the software used are GAUSS and S-PLUS.*

To better visualize the three peaks of the stochastic kernel, in Figure 3 the Chinese regions are clustered into three groups according to their average scores and average changing levels of the CIRD. Here the regions are grouped by the default algorithm of *Natural Breaks*[29] in terms of their grades of average CIRD values and average changes (see Table 2). The three groups benchmarked by the regional CIRD levels are: one in the low socio-economic status, ranged from 0 to 14 of the average CIRD grades; one in the middle level of socio-economic status, ranged from 15 to 38 of the grades; one in the high socio-economic status, ranged from 39 to 100 of the grades. The three groups benchmarked by the regional CIRD changes are: one grows slow in the regional socio-economic development, scored from 0 to 29 in the average changes of CIRD; one grows in an average rate, scored from 30 to 69 in the average changes; one grows fast in the regional socio-economic development, scored from 70 to 100 in the average changes.

[29] In GIS dictionary, the algorithm of *Natural Breaks* classifies data on the basis of "natural groups in the data distribution", which are presented as "the low points of valleys" in the histogram. For example, the first group by *Natural Breaks* is allocated at "the largest valley". (Quoted from the official website of the GIS mapping software applied in the study: http://support.esri.com/en/knowledgebase/GISDictionary/term/natural%20breaks%20classification)

The clustering tendency is largely corresponding to the geographic locations in the map based on the average CIRD scores: the highly developed group includes most of the East coast regions, while the underdeveloped group mainly contains Western inland and peripheral regions, and the rest Central and Northeastern regions constitute the middle level group. In fact, the groups benchmarked by average changes of CIRD show a different picture: the fast-developing group mainly comprises the inland and peripheral regions belonging to the middle or low level groups in the previous classification according to the average CIRD scores, while the slow-developing group contains only 5 members and shows a scattered distribution, i.e., the Northeast, Southwest (Sichuan) and the East coast (Jiangsu); the majority of Chinese regions develops at an average rate. This implies the existence of a convergence process which is however far from been completed.

***Figure 3.*** *Maps: the average level of CIRD and the change of CIRD (1998-2010)*

| *Mapping by regional CIRD levels* | *Mapping by regional CIRD changes* |
|---|---|
|  |  |



*Note: the maps are colored by the PCA scores and drawn by the software ArcView GIS. The map with the names of each region is presented in Appendix 2. (The regions in the color of blank are excluded in the sample because of missing data).*

## 4. Conclusion and Discussion

Contrary to conventional wisdom, the result of this study shows that the dynamic distribution of the development levels across Chinese regions follows a pattern of three-peak stratification. This finding is based on a new multidimensional index (CIRD) that captures all the main pillars of socio-economic development. More importantly, the result challenges the common opinion that the East coast regions or theirs adjacent areas could develop more rapidly along all dimensions.

The analysis shows a tendency of some catching-up in the groups of regions with the lowest levels of development – some inland and peripheral regions have grown relatively faster in certain non GDP-related dimensions. Specifically, some advanced regions adjacent to the coast, such as Hebei, have been overcome by Shanxi and other Central inland and peripheral region (like Inner Mongolia).

The strategy of balancing the uneven growth process in China has not yet been completed: in spite of the remarkable achievements in poverty reduction[30], the regional disparity among the various areas of the country has not been eliminated. The Central and Western regions are still far behind by the East and Coastal areas in various socio-economic aspects. As shown by this analysis, what is particularly important is to tailor specific policies and programs targeting different aspects across the different regions.

Besides its findings, this study raises a number of issues for possible further research. First, it could be interesting to tackle the potential interactions across the five sub-dimensions of the CIRD with a view to seeing whether certain dimensions are positively or negatively correlated with each other. Second, it would be valuable to assess the effects of different policies on regional disparity in China so as to provide more concrete instructions for assessment of the impact of the local policies on different domains. The main challenges to pursue these types of studies lie in the availability of the reliable and consistent data and the small size of the sample under observation. Possible improvements in this direction could be the introduction of a sub-dimensional division of urban-rural disparity or the expansion of the sample to an inferior administrative level of prefectural regions. They both fall beyond the scope of this analysis and, accordingly, remain objects for future research.

[30] It is reported that the proportion of population living in poverty reduced from 81.6% to 10.4%.

***Acknowledgements***

I would like to express my deepest gratitude to Prof. Andrea Fracasso for his selfless helps and valuable contributions to this paper. I also sincerely thank Prof. Luigi Bonatti, Prof. Roberto Fanfani, Prof. Cristina Brasili and Dr. Barbara Barone, for their precious suggestions to this paper. A special acknowledgement shall be made for the funding from the School of International Studies (University of Trento, Italy).

***References***


Barone, Barbara, Peng Bin, Cristina Brasili and Roberto Fanfani, 2013, "Dynamic Changes in China's Spatial Inequalities across Provinces and its Rural and Urban Areas," *L'industria*, Vol. (Issue) 3, pp. 471-490.

Bonatti, Luigi and Andrea Fracasso, 2010, "Global rebalancing and the future of the Sino-US codependency," *China and World Economy,* Vol. 18, pp. 70-87.

Cameron, A. Colin and Pravin K. Trivedi, 2005, *Microeconometrics: Methods and Applications*. Cambridge University Press, Cambridge, UK.

Chen, Bo and Yuen Pau Woo, 2010, "Measuring economic integration in the Asia-Pacific region: A principal components approach," *Asian Economic Papers*, Vol. 9, No. 4, pp. 121–43.

Cheong, Tsun Se and Yanrui Wu, 2013, "Regional disparity, transitional dynamics and convergence in China," *Journal of Asian Economics*, Vol. 29(C), pp. 1-14.

Dreher, Axel, 2006, "Does Globalization Affect Growth? Evidence from a New Index of Globalization," *Applied Economics,* Vol. 38, pp. 1091-1110.

Fan, Shenggen, Ravi Kanbur and Xiaobo Zhang, 2011, "China's regional disparities: Experience and policy*", Review of Development Finance,* Vol. 1, pp. 47-56.

Hardle, Wolfgang, 1990, *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, UK.

Heshmati, Almas, 2006, "Measurement of a multidimensional index of globalization," *Global Economy Journal,* Vol. 6 (2), pp. 1-30.

Hotelling, Harold, 1933, "Analysis of a complex statistical variables into principal components," *Journal of Educational Psychology*, Vol. 24, pp. 417-41.

Hu, Angang, 2003, "One China, four worlds: The inequality of uneven regional development in China", in Hu Angang, Wang Shaoguang, and Zhou Jianming, eds, *China's Second Generation of Reform Ideas: Centered with Institutional Construction*, pp. 1-20. Beijing: Tsinghua University Press (in Chinese).

Li, Hong, 2003, "Dynamics of Income Distribution across Chinese Provinces during 1978-98," *Journal of Chinese Economic and Business Studies*, Vol. 1(2), pp. 145–157.

Liu, Yong and Wei Zou, 2011, "Rural-urban migration and dynamics of income distribution in China: a non-parametric approach," *China and World Economy,* Vol. 19(6), pp. 37–55.

Mazzocchi, Mario, 2008, *Statistics of Marketing and Consumer Research*. London: SAGE Publications.

Mishra, SK, 2007, "A Comparative Study of Various Inclusive Indices and the index Constructed by the Principal Components Analysis", MPRA Paper, No.3377, online at http://mpra.ub.uni-muenchen.de/3377/

OECD, 2008, *Handbook on constructing composite indicators: Methodology and user guide*. Paris: OECD Publisher.

Quah, Danny T., 1993, "Galton's Fallacy and the Tests of Convergence Hypothesis". *Scandinavia Journal of Economics*, Vol. 95(4), pp. 427–433.

Quah, Danny T., 1995, "Empirics for economic growth and convergence", Centre for Economic Performance, Discussion paper, No. 253. Online at http://cep.lse.ac.uk/pubs/download/DP0253.pdf

Quah, Danny T., 1996, "Twin peaks: Growth and convergence in models of distribution dynamics", *Economic Journal*, Vol. 106, pp. 1045–1055.

Quah, Danny T., 1997, "Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs", *Journal of Economic Growth*, Vol. 2, Issue 1, pp. 27-59.

Sakamoto, Hiroshi and Jin Fan, 2010, "Distribution dynamics and convergence among 75 cities and counties in Yangtze River delta in China: 1990–2005", *Review of Urban and Regional Development Studies*, Vol. 22, pp.39–54.

Sakamoto, Hiroshi and Nazrul Islam, 2008, "Convergence across Chinese provinces: an analysis using Markov transition matrix", *China Economic Review*, Vol. 19, pp. 66–79.

Saisana, Michaela and Stefano Tarantola, 2002, "State-of-the-art report on current methodologies and practices for composite indicator development," EUR 20408 EN. European Commission, Joint Research Centre -JRC: Italy.

Smith, Lindsay I., 2002, A Tutorial on Principal Component Analysis. Online at: http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

Villaverde Jose, Maza Adolfo, and Ramasamy Bala, 2010, "Provincial disparities in post-reform China". *China and World Economy*, Vol. 18, pp. 73–95.

Wei, Y.H. Dennis and Xinyue Ye, 2009, "Beyond convergence: Space, scale, and regional inequality in China". *Tijdschrift voor Economische en Sociale Geografie*, Vol. 100, No. 1, pp. 59–80.

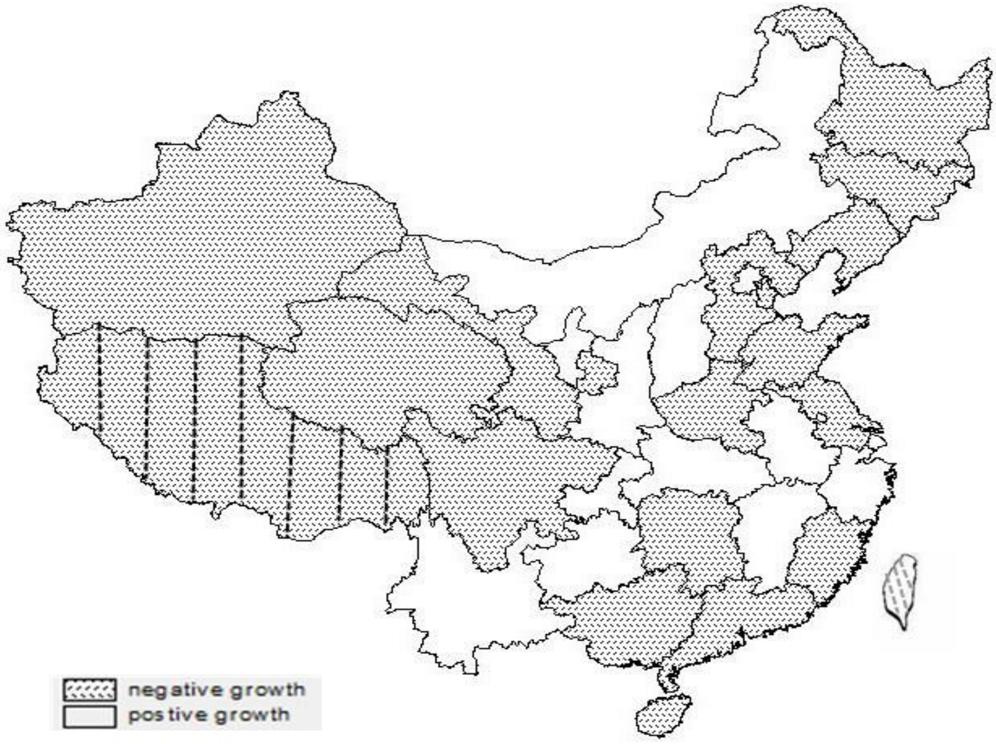
Urban China Initiative (UCI), 2010, "The urban sustainability index: A new tool for measuring China's cities," Report, Online at http://www.mckinsey.it/storage/first/uploadfile/attach/142331/file/the_urban_sustainability_index.pdf

Zambom, Adriano Zanin and Ronaldo Dias, 2013, "A review of Kernel density estimation with applications to econometrics," *International Econometric Review*, Vol. 5, pp. 20–42.

Zhou, Hao and Wei Zou, 2010, "Income distribution dynamics of urban residents: The case of China (1995–2004)," *Frontiers of Economics in China*, Vol. 5, pp. 114–134.

Zou, Wei and Hao Zhou, 2007, "Classification of growth clubs and convergence: evidence from panel data in China, 1981–2004", *China and World Economy*, Vol. 15, No. 5, pp. 91–106.

Appendix 1. The regions with negative growth in 1998-2010




Appendix 2. Chinese Map of Administrative Districts

Appendix 3. The Rankings of Chinese regions in terms of the Min-Max scores of CIRD (1998-2010)

| Regions | Average CIRD | Regions | Average changing levels |
|---|---:|---|---:|
| Beijing | 4,95 | Beijing | 0,02 |
| Shanghai | 4,48 | Zhejiang | 0,02 |
| Jiangsu | 3,55 | Ningxia | 0,02 |
| Shandong | 3,40 | Jiangxi | 0,02 |
| Zhejiang | 3,36 | Shaanxi | 0,01 |
| Tianjin | 3,35 | Inner Mongolia | 0,01 |
| Guangdong | 3,26 | Shanghai | 0,01 |
| Fujian | 2,69 | Shanxi | 0,01 |
| Liaoning | 2,67 | Guizhou | 0,01 |
| Hubei | 2,42 | Anhui | 0,01 |
| Hebei | 2,18 | Chongqing | 0,00 |
| Henan | 2,16 | Hubei | 0,00 |
| Heilongjiang | 2,12 | Yunnan | 0,00 |
| Hunan | 2,10 | Henan | -0,01 |
| Hainan | 2,10 | Guangxi | -0,01 |
| Chongqing | 2,07 | Hebei | -0,01 |
| Jilin | 2,06 | Guangdong | -0,01 |
| Shanxi | 2,05 | Gansu | -0,01 |
| Anhui | 2,01 | Shandong | -0,02 |
| Sichuan | 1,99 | Qinghai | -0,02 |
| Shaanxi | 1,98 | Hunan | -0,02 |
| Yunnan | 1,79 | Xinjiang | -0,02 |
| Jiangxi | 1,79 | Tianjin | -0,02 |
| Guangxi | 1,77 | Fujian | -0,03 |
| Inner Mongolia | 1,65 | Sichuan | -0,03 |
| Xinjiang | 1,59 | Jiangsu | -0,03 |
| Gansu | 1,56 | Jilin | -0,04 |
| Ningxia | 1,50 | Hainan | -0,04 |
| Guizhou | 1,40 | Liaoning | -0,04 |
| Qinghai | 1,28 | Heilongjiang | -0,05 |